



501.43507X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): KAIYA, et al

Serial No.: 10/774,470

Filed: February 10, 2004

Title: EXTERNAL STORAGE AND RECOVERY METHOD FOR EXTERNAL
STORAGE AS WELL AS PROGRAM

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 9, 2004

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby
claim(s) the right of priority based on:

Japanese Patent Application No. 2003-076865
Filed: March 20, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl L. Brundidge
Registration No.: 29,621

CIB/rr
Attachment

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年 3月20日

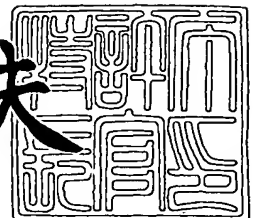
出願番号
Application Number: 特願2003-076865
[ST. 10/C]: [JP 2003-076865]

願人
Applicant(s): 株式会社日立製作所

2004年 2月19日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2004-3011053

【書類名】 特許願
【整理番号】 340201769
【あて先】 特許庁長官殿
【国際特許分類】 G06F 03/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 3 2 番地 2 号 株式会社日立製作所
RAIDシステム事業部内

【氏名】 海谷 佳一

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 3 2 番地 2 号 株式会社日立製作所
RAIDシステム事業部内

【氏名】 坪木 雅直

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 3 2 番地 2 号 株式会社日立製作所
RAIDシステム事業部内

【氏名】 水主 和人

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100095371

【弁理士】

【氏名又は名称】 上村 輝之

【選任した代理人】

【識別番号】 100089277

【弁理士】

【氏名又は名称】 宮川 長夫

【選任した代理人】

【識別番号】 100104891

【弁理士】

【氏名又は名称】 中村 猛

【手数料の表示】

【予納台帳番号】 043557

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0110323

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラム

【特許請求の範囲】

【請求項 1】 ホストコンピュータに接続される外部記憶装置であつて、
前記ホストコンピュータにより利用されるデータを記憶する記憶手段と、
前記記憶手段を制御する制御手段とを有し、
前記制御手段は、

前記記憶手段に記憶されたデータに関して前記ホストコンピュータにより設定される回復可能時点を登録する登録手段と、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、
を備えたことを特徴とする外部記憶装置。

【請求項 2】

前記登録手段は、前記ホストコンピュータにより設定される任意の複数時点を、前記回復可能時点として登録可能である請求項 1 に記載の外部記憶装置。

【請求項 3】

前記記憶手段は、前記ホストコンピュータからの書込みデータをジャーナルデータとして記憶するジャーナルデータ記憶手段を有し、

前記登録手段は、前記ホストコンピュータからの指示に基づいて、前記ジャーナルデータの所定位置に標識情報を対応付けることにより、前記回復可能時点を登録するものである請求項 1 に記載の外部記憶装置。

【請求項 4】

前記ジャーナルデータは、少なくとも、書込みデータと、書込み位置と、前記標識情報としての回復フラグ情報とを含んで構成され、

前記登録手段は、前記ジャーナルデータ中の所定の回復フラグ情報をセットすることにより、前記回復可能時点を登録するものである請求項 3 に記載の外部記

憶装置。

【請求項 5】

前記記憶手段は、バックアップデータを記憶するバックアップデータ記憶手段を有し、

前記制御手段は、ジャーナルデータ管理手段を有し、

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶手段の空き容量が不足した場合には、前記ジャーナルデータ記憶手段に記憶されている最古のジャーナルデータを前記バックアップデータ記憶手段に移し替えて、前記ジャーナルデータ記憶手段の空き容量を増加させ、かつ、前記登録された回復可能時点のうち最古の回復可能時点が変更された旨を前記ホストコンピュータに通知するものである請求項 3 に記載の外部記憶装置。

【請求項 6】

前記制御手段は、ジャーナルデータ管理手段を有し、

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶手段の空き容量が不足した場合には、前記記憶手段内の未使用の記憶領域を利用してジャーナルデータ記憶手段の論理サイズを自動的に拡張させるものである請求項 3 に記載の外部記憶装置。

【請求項 7】 ホストコンピュータに接続された外部記憶装置のデータを、該外部記憶装置内で回復させるデータ回復方法であって、

記憶されたデータに関して前記ホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録ステップと、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する一覧送信ステップと、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復ステップと、
を含んだことを特徴とする外部記憶装置のデータ回復方法。

【請求項 8】 ホストコンピュータに接続された外部記憶装置を制御するためのプログラムであって、

前記外部記憶装置は、前記ホストコンピュータにより利用されるデータを記憶

する記憶手段を有し、

前記記憶手段に記憶されたデータに関して前記ホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録手段と、

前記ホストコンピュータからの要求に応じて、前記登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、

前記回復可能時点の選択用情報に基づいて前記ホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、

を、外部記憶装置のコンピュータ上に実現させるプログラム。

【請求項 9】

ジャーナルデータを取得して前記記憶手段のジャーナルデータ記憶領域に記憶させるジャーナルデータ管理手段を、前記外部記憶装置のコンピュータ上に実現させると共に、

前記登録手段は、前記ホストコンピュータからの指示に基づいて、前記ジャーナルデータの所定位置に標識情報を対応付けることにより、前記回復可能時点を登録するものである請求項 8 に記載のプログラム。

【請求項 10】

前記ジャーナルデータ管理手段は、前記ジャーナルデータ記憶領域の空き容量が不足した場合には、前記ジャーナルデータ記憶領域に記憶されている最古のジャーナルデータを前記記憶手段のバックアップデータ記憶領域に移し替えて、前記ジャーナルデータ記憶領域の空き容量を増加させ、前記登録された回復可能時点のうち最古の回復可能時点が変更された旨を前記ホストコンピュータに通知するものである請求項 9 に記載のプログラム。

【請求項 11】 外部記憶装置を利用するホストコンピュータを制御するプログラムであって、

前記外部記憶装置に記憶されたデータに関して、任意の複数時点で設定可能な回復可能時点を前記外部記憶装置に指示し登録させる登録指示手段と、

前記外部記憶装置に登録された前記回復可能時点の選択用情報を要求する選択用情報要求手段と、

前記外部記憶装置から受信した前記選択用情報に基づいて、所望のデータを所

望の回復可能時点まで回復させるように前記外部記憶装置に指示する回復指示手段と、

をホストコンピュータ上で実現させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、例えば、ディスク装置等の外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラムに関する。

【0002】

【従来の技術】

比較的多量のデータを取り扱う業務用アプリケーションプログラム（データベースシステム）では、ホストコンピュータとは別体に形成されたディスクアレイ装置にデータを保存する。そして、ホストコンピュータのデータベースシステムからディスクアレイ装置上のデータにアクセスして、種々のデータ操作を行うようになっている。ディスクアレイ装置とは、複数のディスク装置をアレイ状に配設してなるもので、ホストコンピュータからの書込み命令や読出し命令等に応じて作動するようになっている。

【0003】

ここで、データベースシステムが稼働中に、例えば、予期せぬ電源切断、オペレータの操作ミス、ハードウェア回路や他のプログラムの不調等によって、障害が発生した場合、データベースの内容を障害発生前の状態に復旧させる必要を生じる。また、障害以外に所望の時点までデータ操作を戻したい場合もある。

【0004】

第1の従来技術として、通常のデータベースシステムでは、ホストコンピュータ上のデータベースシステム自身が、実データとは別にジャーナルデータ（ログデータ）をディスクアレイ装置の所定のディスク装置に書き出すようになっている。従って、通常のデータベースシステムでは、事前に取得してあるバックアップデータを基に、データベースシステム自身がディスク装置からジャーナルデータを読み出して、バックアップデータに順次反映させていく。これにより、ホス

トコンピュータ上のデータベースシステムは、ジャーナルデータが残存している範囲内で、所望の時点にデータベースを復旧させることができる。

【0005】

第2の従来技術では、第1のディスク装置の内容を所定の周期でバックアップ用ディスク装置に保存すると共に、ジャーナルデータをジャーナル用ディスク装置に保存する。第1のディスク装置に障害が発生した場合は、バックアップデータ及びジャーナルデータに基づいて、第2のディスク装置内に仮想的な第1のディスク装置を生成し、第1のディスク装置へのデータアクセスを、仮想的な第1のディスク装置に内部的に切り替える。そして、第1のディスク装置の修復が完了すると、仮想的な第1のディスク装置の内容を第1のディスク装置に移し替えるようになっている（例えば、特許文献1参照）。

【0006】

【特許文献1】

特開平6-110618号公報

【0007】

【発明が解決しようとする課題】

上記第1の従来技術では、ホストコンピュータ上のデータベースシステム自身がジャーナルデータを管理しており、任意の時点にデータを復旧可能である。しかし、データベースシステム自身がデータ復旧作業を行うため、ホストコンピュータのコンピュータ資源（演算ユニットやメモリ等）が、データ復旧処理に使用されてしまい、復旧作業中に、本来の業務処理や他の業務処理の処理効率を低下させることになる。また、データベースシステムがジャーナルデータの管理を行っているが、ジャーナルデータの格納ディスクが満杯になってしまうと、バックアップデータを取っていない限り、データを復旧させることはできない。従って、データベースシステムは、ジャーナルデータ用ディスクの容量管理等まで行う必要があり、処理負担が大きくなる。さらに、データの世代管理を行う場合は、複数世代のバックアップデータを作成するため、処理負担が更に増大する。

【0008】

第2の従来技術では、仮想的なディスク装置にアクセスを切り替えることによ

り、実行中の処理を中断することなく、データ復旧作業を行うことができるが、直前の状態までしか回復させることができず、オペレータが所望する任意の時点にデータを復旧させることができない。

【0009】

本発明は、上記問題点に鑑みてなされたもので、その目的は、ホストコンピュータ側の処理負担を増大させずに、任意の時点へデータを復旧させることができる外部記憶装置及び外部記憶装置のデータ回復方法並びにプログラムを提供することにある。本発明のさらなる目的は、後述する実施の形態の記載から明らかになるであろう。

【0010】

【課題を解決するための手段】

上記課題を解決すべく、本発明の第1の観点に従う外部記憶装置は、ホストコンピュータに接続されるもので、ホストコンピュータにより利用されるデータを記憶する記憶手段と、記憶手段を制御する制御手段とを有する。制御手段は、記憶手段に記憶されたデータに関してホストコンピュータにより設定される回復可能時点を登録する登録手段と、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報をホストコンピュータに送信する選択用情報送信手段と、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段と、を備える。

【0011】

記憶手段としては、例えば、複数のディスク装置をアレイ状に配置してなる記憶装置を用いることができる。ホストコンピュータは、記憶手段に記憶されるデータに関して、回復可能時点を設定することができる。回復可能時点とは、そのデータを回復させることが可能な時点を示す情報であり、復元ポイントと称することもできる。ホストコンピュータにより定期的に又は不定期に設定される回復可能時点は、登録手段により登録される。

【0012】

障害が発生等してデータの復旧が必要になった場合、ホストコンピュータは、制御手段に対して、回復可能時点の選択用情報を要求する。この要求に応じて、

選択用情報送信手段は、選択用情報をホストコンピュータに送信する。選択用情報とは、回復可能時点を選択するための情報であり、例えば、一覧形式等に表示可能である。

【0013】

ホストコンピュータは、受信した選択用情報に基づいて、回復させたい時点を選択する。ホストコンピュータにより選択された回復可能時点は、回復手段に通知される。そして、回復手段は、ホストコンピュータに指定されたデータを指定された時点まで回復させる。回復手段は、例えば、指定された回復時点までのジャーナルデータをバックアップデータに順次反映させていくことにより、データを復旧させることができる。これにより、ホストコンピュータのコンピュータ資源を事実上殆ど使用することなく、外部記憶装置内で任意の時点までデータを回復させることができる。

【0014】

登録手段は、ホストコンピュータにより設定される任意の複数時点を、回復可能時点として登録可能である。即ち、直前の最新状態のみならず、任意の時点を複数個登録させることができる。例えば、ホストコンピュータは、更新処理（コミット）を要求する度毎に、あるいは、データ操作の区切りがつかう度毎に、自動的に又はオペレータによる手動で、回復可能時点を設定することができる。

【0015】

本発明の一態様では、記憶手段は、ジャーナルデータを取得して記憶するジャーナルデータ記憶手段を有し、登録手段は、ホストコンピュータからの指示に基づいて、ジャーナルデータの所定位置に標識情報を対応付けることにより、回復可能時点を登録するようになっている。即ち、ジャーナルデータは、外部記憶装置内のジャーナルデータ記憶手段が、独自に自動的に収集して記憶する。そして、登録手段は、ホストコンピュータからの設定に基づいて、ジャーナルデータの所定位置に標識情報を対応付けることにより、回復可能時点を登録する。標識情報は、ジャーナルデータ中に含めることもできるし、ジャーナルデータとは別のデータとしてそれぞれ管理し、ユニークな識別コード等で両者を関連付けることもできる。

【0016】

本発明の一態様では、ジャーナルデータは、少なくとも、書込みデータと、書込み位置と、標識情報としての回復フラグ情報とを含んでなり、登録手段は、ジャーナルデータ中の所定の回復フラグ情報をセットすることにより、回復可能時点を登録するようになっている。

【0017】

回復フラグを追加してジャーナルデータのデータ構造を拡張する。全てのジャーナルデータには回復フラグをセットするデータ領域が予め含まれている。あるデータについて回復可能時点を設定する場合は、該データに対応する回復フラグをセットする。回復フラグをリセットすれば、設定された回復可能時点を解除することができる。

【0018】

本発明の一態様では、さらに、記憶手段は、バックアップデータを記憶するバックアップデータ記憶手段を有し、制御手段は、ジャーナルデータ管理手段を有している。そして、ジャーナルデータ管理手段は、ジャーナルデータ記憶手段の空き容量が不足した場合には、ジャーナルデータ記憶手段に記憶されている最古のジャーナルデータをバックアップデータ記憶手段に移し替えて、ジャーナルデータ記憶手段の空き容量を増加させ、かつ、登録された回復可能時点のうち最古の回復可能時点が変更された旨をホストコンピュータに通知する。

【0019】

データの回復は、例えば、ある時点のバックアップデータに、目標とする時点までのジャーナルデータを順次反映させることにより実現される（ロールフォワード方式）。従って、ジャーナルデータが存在しない場合は、バックアップされた時点にしかデータを戻すことはできない。一方、ジャーナルデータは、データ更新履歴の集合体であり、日々増大する。ジャーナルデータの保存量が、ディスク装置の記憶容量に達すると、それ以上のジャーナルデータを記憶することはできない。そこで、ジャーナルデータの空き容量が不足した場合は、既に蓄積されているジャーナルデータの中から最も古いデータを必要な量だけバックアップデータに移し替えて、空き容量を確保する。移し替える必要量は、予め設定され

た固定値としても良いし、ジャーナルデータの蓄積速度やバックアップデータ記憶手段の記憶容量等の諸要因に応じて動的に変化させてもよい。ここで、最古のジャーナルデータをバックアップデータに移し替えるとは、最古のジャーナルデータをバックアップデータに反映させた上で、最古のジャーナルデータを削除することを意味する。なお、記憶手段内に未使用の記憶領域がある限り、ジャーナルデータ記憶領域を自動的に拡張させ、未使用の記憶領域が不足した場合に、最古のジャーナルデータをバックアップデータに移し替えるようにしてもよい。

【0020】

本発明の第2の観点に従う外部記憶装置のデータ回復方法は、ホストコンピュータに接続された外部記憶装置のデータを、該外部記憶装置内で回復させるデータ回復方法であって、記憶されたデータに関してホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録ステップと、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報をホストコンピュータに送信する一覧送信ステップと、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復ステップと、を含んだことを特徴とする。

【0021】

登録ステップ、一覧送信ステップ、回復ステップは、この順序で実行してもよいし、異なる順序、例えば、並行的に実行してもよい。

【0022】

本発明の第3の観点に従うプログラムは、ホストコンピュータに接続された外部記憶装置を制御するためのプログラムであって、外部記憶装置は、ホストコンピュータにより利用されるデータを記憶する記憶手段を有し、記憶手段に記憶されたデータに関してホストコンピュータにより任意の複数時点に設定されうる回復可能時点を登録する登録手段と、ホストコンピュータからの要求に応じて、登録された回復可能時点の選択用情報を前記ホストコンピュータに送信する選択用情報送信手段と、回復可能時点の選択用情報に基づいてホストコンピュータから指定されたデータを指定された回復可能時点まで回復させる回復手段とを外部記憶装置のコンピュータ上に実現させる。

【0023】

本発明の第4の観点に従うプログラムは、外部記憶装置を利用するホストコンピュータを制御するプログラムであって、外部記憶装置に記憶されたデータに関して、任意の複数時点で設定可能な回復可能時点を外部記憶装置に指示し登録させる登録指示手段と、外部記憶装置に登録された回復可能時点の選択用情報を要求する選択用情報要求手段と、外部記憶装置から受信した選択用情報に基づいて、所望のデータを所望の回復可能時点まで回復させるように外部記憶装置に指示する回復指示手段と、をホストコンピュータ上で実現させる。

【0024】

このプログラムは、例えば、API (Application Program Interface) のような形で提供可能であり、種々の業務用アプリケーションプログラムから好適に利用されることができる。

【0025】

本発明に従うプログラムは、例えば、ディスク型記憶媒体、半導体メモリ等の各種記憶媒体に固定して流通に置くこともできるし、あるいは、サーバから通信ネットワークを介して配信することもできる。

【0026】

【発明の実施の形態】

以下、図1～図10に基づき、本発明の実施の形態を説明する。

【0027】

まず最初に、図1に基づいて、外部記憶システムの全体概要を説明する。

【0028】

先にシステムの全体構成を図1に基づいて説明する。記憶装置システム60は、記憶デバイス制御装置10と記憶デバイス30とを備えて構成されている。記憶デバイス制御装置10は、情報処理装置20から受信したコマンドに従って、記憶デバイス30に対する制御を行う。例えば、記憶デバイス制御装置10は、情報処理装置20からデータの入出力要求を受信すると、記憶デバイス30に記憶されているデータの入出力処理を行う。記憶デバイス30が備えるディスクドライブにより提供される物理的な記憶領域上には、論理ボリューム (Logical Un

it) (以下、LUと略記) が設定されている。LUは、論理的な記憶領域であり、このLU上にデータは記憶されている。また、記憶デバイス制御装置10は、情報処理装置20との間で、記憶装置システム60を管理するための各種コマンドの授受も行う。

【0029】

情報処理装置20は、CPU (Central Processing Unit) やメモリ等を備えたコンピュータシステムである。情報処理装置20のCPUが各種プログラムを実行することにより、種々の機能が実現される。情報処理装置20は、例えば、パーソナルコンピュータやワークステーションである場合もあるし、メインフレームコンピュータの場合もある。図1では、説明の便宜上、第1～第5の5台の情報処理装置を図示する。各情報処理装置20を識別するために、図1中では「情報処理装置1」、「情報処理装置2」等のように連番を付し、第1～第5の情報処理装置20とする。後述のチャンネル制御部11及びディスク制御部14も同様に連番を付して区別する。

【0030】

第1～第3の情報処理装置20は、LAN (Local Area Network) 40を介して、記憶デバイス制御装置10と接続されている。LAN40は、例えば、インターネットとすることもできるし、専用のネットワークとすることもできる。第1～第3の情報処理装置20と記憶デバイス制御装置10との間のデータ通信は、LAN40を介して、例えば、TCP/IP (Transmission Control Protocol/Internet Protocol) プロトコルに従って行われる。第1～第3の情報処理装置20からは、記憶装置システム60に対して、ファイル名指定によるデータアクセス要求 (ファイル単位でのデータ入出力要求である。以下、「ファイルアクセス要求」と略記) が送信される。

【0031】

LAN40には、バックアップデバイス71が接続されている。バックアップデバイス91としては、例えば、MO (magneto-optic: 光磁気型記憶装置)、CD-R (CD-Recordable: 読み書き可能なコンパクトディスク)、DVD-RAM (Digital Versatile Disk-RAM: 読み書き可能なDVD) 等のディスク系記憶デバイス

や、例えば、DAT (Digital Audio Tape) テープ、カセットテープ、オープンテープ、カートリッジテープ等のテープ系記憶デバイスを用いることができる。バックアップデバイス 71 は、LAN 40 を介して記憶デバイス制御装置 10 との間で通信を行うことにより、記憶デバイス 30 に記憶されているデータのバックアップデータを記憶する。また、バックアップデバイス 71 は、第 1 の情報処理装置 20 と接続されるように構成することもできる。この場合は、第 1 の情報処理装置 20 を介して、記憶デバイス 30 に記憶されているデータのバックアップデータを取得するようにする。

【0032】

記憶デバイス制御装置 10 は、第 1～第 4 のチャンネル制御部 11 により、LAN 40 を介して第 1～第 3 の情報処理装置 20 やバックアップデバイス 71 との間で通信を行う。第 1～第 4 のチャンネル制御部 11 は、第 1～第 3 の情報処理装置 20 からのファイルアクセス要求を個々に受け付ける。即ち、第 1～第 4 のチャンネル制御部 11 には、それぞれ LAN 40 上のネットワークアドレス（例えば、IP アドレス）が割り当てられており、第 1～第 4 の各チャンネル制御部 11 はそれぞれが個別に NAS (Network Attached Storage) として振る舞うようになっている。第 1～第 4 のチャンネル制御部 11 は、それぞれが独立した NAS であるかのように、第 1～第 3 の情報処理装置 20 に対し NAS としてのサービスを提供可能である。以下、第 1～第 4 のチャンネル制御部 11 を CHN と略す場合がある。このように、1 台の記憶装置システム 60 内にそれぞれ個別に NAS としてのサービスを提供する第 1～第 4 のチャンネル制御部 11 を備えるように構成したことにより、従来、独立したコンピュータで個々に運用されていた NAS サーバが 1 台の記憶装置システム 60 に集約される。そして、これにより、記憶装置システム 60 の統括的な運用が可能となり、各種設定・制御や障害管理、バージョン管理といった保守業務の効率化を図ることができる。

【0033】

なお、記憶デバイス制御装置 10 の第 1～第 4 のチャンネル制御部 11 は、例えば、一体的にユニット化された回路基板上に形成されたハードウェア、このハードウェアにより実行される OS (Operating System)、この OS 上で動作するア

アプリケーションプログラム等のソフトウェアにより実現される。記憶装置システム 60 では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。従って、記憶装置システム 60 を用いることにより、柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズにきめ細やかに対応可能となる。

【0034】

第3及び第4の情報処理装置 20 は、SAN (Storage Area Network) 50 を介して、記憶デバイス制御装置 10 と接続されている。SAN 50 は、記憶デバイス 30 が提供する記憶領域におけるデータの管理単位であるブロックを単位として、第3及び第4の情報処理装置 20 との間でデータの授受を行うためのネットワークである。SAN 50 を介して行われる第3及び第4の情報処理装置 20 と記憶デバイス制御部 10 との間の通信は、一般にファイバチャネルプロトコルに従う。第3及び第4の情報処理装置 20 からは、記憶装置システム 60 に対して、ファイバチャネルプロトコルに従ってブロック単位でのデータアクセス要求（以下、ブロックアクセス要求と略記）が送信される。

【0035】

SAN 50 には、SAN 対応のバックアップデバイス 70 が接続されている。SAN 対応バックアップデバイス 70 は、SAN 50 を介して記憶デバイス制御装置 10 との間で通信を行うことにより、記憶デバイス 30 に記憶されているデータのバックアップデータを記憶する。

【0036】

記憶デバイス制御装置 10 は、第5及び第6のチャネル制御部 11 により、SAN 50 を介して第3及び第4の情報処理装置 20 及び SAN 対応バックアップデバイス 70 との間の通信を行う。以下、第5及び第6のチャネル制御部 11 を CHF と略記する場合がある。

【0037】

また、第5の情報処理装置 20 は、LAN 40 や SAN 50 等のネットワークを介さずに、記憶デバイス制御装置 10 と直接的に接続されている。第5の情報処理装置 20 としては、例えば、メインフレームコンピュータとすることができ

るが、もちろんこれに限定されない。第5の情報処理装置20と記憶デバイス制御装置10との間の通信は、例えば、FICON (Fibre Connection) (登録商標) やESCON (Enterprise System Connection) (登録商標)、ACONARC (Advanced Connection Architecture) (登録商標)、FIBARC (Fibre Connection Architecture) (登録商標) 等の通信プロトコルに従う。第5の情報処理装置20からは、記憶装置システム60に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

【0038】

記憶デバイス制御装置10は、第7及び第8のチャネル制御部11により、第5の情報処理装置20との間で通信を行う。以下、第7及び第8のチャネル制御部11をCHAと略す場合がある。

【0039】

SAN50には、記憶装置システム60の設置場所（プライマリサイト）から遠隔した場所（セカンダリサイト）に設置される他の記憶装置システム61が接続されている。他の記憶装置システム61は、レプリケーションまたはリモートコピーの機能におけるデータ複製先の装置として利用される。なお、他の記憶装置システム61は、SAN50以外にも例えばATM (Asynchronous Transfer Mode) 等の通信回線を介して記憶装置システム60に接続される場合もある。この場合には、上記通信回線を利用するためのインターフェース（チャネルエクステンダ）を備えたチャネル制御部11が採用される。

【0040】

次に、記憶デバイス30の構成について説明する。記憶デバイス30は、多数のディスクドライブ（物理ディスク）を備えており、情報処理装置20に対して記憶領域を提供する。データは、論理的記憶領域であるLUに記憶されている。ディスクドライブとしては、例えば、ハードディスク装置、フレキシブルディスク装置、半導体記憶装置等の種々のデバイスを用いることができる。なお、記憶デバイス30は、例えば、複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置20に対しては、RAID (Redundant Array of Independent (Inexpensive) Disks) により管理された

複数のディスクドライブにより記憶領域を提供することができる。

【0041】

記憶デバイス制御装置 10 と記憶デバイス 30 とは、図 1 に示すように、直接的に接続してもよいし、ネットワークを介して間接的に接続するようにしてもよい。さらに、記憶デバイス 30 は、記憶デバイス制御装置 10 と一体のものとして構成することもできる。

【0042】

記憶デバイス 30 に設定される LU には、情報処理装置 20 からアクセス可能なユーザ LU や、チャンネル制御部 11 の制御のために使用されるシステム LU 等がある。システム LU には、CHN 11 で実行される OS も格納される。また、各 LU には、各チャンネル制御部 11 が予め対応付けられている。これにより、各チャンネル制御部 11 毎にアクセス可能な LU がそれぞれ割り当てられている。また、上記対応付けは、複数のチャンネル制御部 11 で一つの LU を共有するように設定することもできる。なお、以下の説明において、ユーザ LU をユーザディスク、システム LU をシステムディスクと記す場合がある。また、複数のチャンネル制御部 11 により共有される LU を、共有 LU または共有ディスクと記す場合がある。

【0043】

次に、記憶デバイス制御装置 10 の構成を説明する。記憶デバイス制御装置 10 は、チャンネル制御部 11、共有メモリ 12、キャッシュメモリ 13、ディスク制御部 14、接続部 15 及び管理端末 16 を備えている。

【0044】

チャンネル制御部 11 は、情報処理装置 20 との間で通信を行うための通信インターフェースを有し、情報処理装置 20 との間でデータ入出力コマンド等を授受する機能を備えている。例えば、CHN 11 は、第 1～第 3 の情報処理装置 20 からのファイルアクセス要求を受け付ける。これにより、記憶装置システム 60 は、NAS としてのサービスを第 1～第 3 の情報処理装置 20 に提供することができる。また、CHF 11 は、第 3 及び第 4 の情報処理装置 20 からのファイバチャンネルプロトコルに従ったブロックアクセス要求を受け付ける。これにより、

記憶装置システム 60 は、高速アクセス可能なデータ記憶サービスを第 3 及び第 4 の情報処理装置 20 に対して提供することができる。また、CHA11 は、第 5 の情報処理装置 20 からの FICON や ESCON、ACONARC、FIBARC 等のプロトコルに従ったブロックアクセス要求を受け付ける。これにより、記憶装置システム 60 は、第 5 の情報処理装置 20 のようなメインフレームコンピュータ等に対してもデータ記憶サービスを提供することができる。

【0045】

各チャネル制御部 11 は、管理端末 16 と共に内部 LAN 17 で接続されている。これにより、チャネル制御部 11 に実行させるプログラム等を、管理端末 16 からチャネル制御部 11 に送信してインストールさせることも可能となっている。チャネル制御部 11 の構成については、さらに後述する。

【0046】

接続部 15 は、各チャネル制御部 11、共有メモリ 12、キャッシュメモリ 13、各ディスク制御部 14 を相互に接続する。チャネル制御部 11、共有メモリ 12、キャッシュメモリ 13 及びディスク制御部 14 間でのデータやコマンドの授受は、接続部 15 を介することにより行われる。接続部 15 は、例えば、高速スイッチングによりデータ伝送を行う超高速クロスバススイッチ等の高速バスで構成される。チャネル制御部 11 同士を高速バスで接続することにより、個々のコンピュータ上で動作する NAS サーバを LAN を介して接続する場合よりも、チャネル制御部 11 間の通信パフォーマンスが向上する。また、これにより、高速なファイル共有機能や高速フェイルオーバー等が可能となる。

【0047】

共有メモリ 12 及びキャッシュメモリ 13 は、各チャネル制御部 11 及び各ディスク制御部 14 により共有される記憶メモリである。共有メモリ 12 は、主に制御情報やコマンド等を記憶するために利用される。キャッシュメモリ 13 は、主にデータを記憶するために利用される。

【0048】

例えば、あるチャネル制御部 11 が情報処理装置 20 から受信したデータ入出力コマンドが書込みコマンドであった場合、当該チャネル制御部 11 は、書込み

コマンドを共有メモリ 12 に書き込むと共に、情報処理装置 20 から受信した書き込みデータをキャッシュメモリ 13 に書き込む。一方、ディスク制御部 14 は、共有メモリ 12 を監視している。ディスク制御部 14 は、共有メモリ 12 に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ 13 から書き込みデータを読み出し、読み出したデータを記憶デバイス 30 に書き込む。

【0049】

ディスク制御部 14 は、記憶デバイス 30 の制御を行う。例えば、上述のように、ディスク制御部 14 は、チャンネル制御部 11 が情報処理装置 20 から受信した書き込みコマンドに従って、記憶デバイス 30 へデータの書き込みを行う。また、ディスク制御部 14 は、チャンネル制御部 11 から送信された論理アドレス指定による LU へのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。ディスク制御部 14 は、記憶デバイス 30 における物理ディスクが RAID により管理されている場合は、RAID 構成に従ったデータのアクセスを行う。また、ディスク制御部 14 は、記憶デバイス 30 に記憶されたデータの複製管理の制御及びバックアップ制御も行う。さらに、ディスク制御部 14 は、災害発生時のデータ消失防止（ディザスタリカバリ）等を目的として、プライマリサイトの記憶装置システム 60 のデータの複製をセカンダリサイトに設置された他の記憶装置システム 61 にも記憶させる制御（レプリケーション機能またはリモートコピー機能と呼ばれる）等も行う。

【0050】

各ディスク制御部 14 は、管理端末 16 と共に内部 LAN 17 を介して接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部 14 に実行させるプログラム等を、管理端末 16 からディスク制御部 14 に送信してインストールさせることが可能となっている。

【0051】

次に、管理端末 16 について説明する。管理端末 16 は、記憶装置システム 60 を保守・管理するためのコンピュータである。管理端末 16 を操作することにより、例えば、記憶デバイス 30 内の物理ディスク構成の設定や、LU の設定、

チャンネル制御部 11 で実行させるためのプログラムのインストール等を行うことができる。ここで、記憶デバイス 30 内の物理ディスク構成の設定としては、例えば、物理ディスクの増設や減設、RAID 構成の変更（RAID 1 から RAID 5 への変更等）などを挙げることができる。さらに、管理端末 16 からは、記憶装置システム 60 の動作状態の確認や故障部位の特定、チャンネル制御部 11 で実行される OS のインストール等の作業を行うこともできる。また、管理端末 16 は、LAN や電話回線等で外部保守センタと接続されており、外部保守センタから管理端末 16 を利用して記憶装置システム 60 の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は、例えば、OS やアプリケーションプログラム、ドライバソフトウェア等から通知される。この通知は、例えば、HTTP (HyperText Transfer Protocol) プロトコルや SNMP (Simple Network Management Protocol) プロトコル、電子メール等により行うことができる。これらの設定や制御は、管理端末 16 上で動作するウェブサーバが提供するウェブページをユーザインターフェースとして、オペレータ等が操作することにより行うことができる。オペレータ等は、管理端末 16 を操作して、障害監視の対象や内容を設定したり、障害通知先を設定等する。

【0052】

管理端末 16 は、記憶デバイス制御装置 10 内に内蔵させる構成でもよいし、記憶デバイス制御装置 10 に外付けする構成でもよい。また、管理端末 16 は、記憶デバイス制御装置 10 及び記憶デバイス 30 の保守・管理を専用に行うコンピュータとして構成してもよいし、あるいは、汎用コンピュータに保守・管理機能を持たせることにより構成してもよい。

【0053】

次に、図 2 を参照して本発明によるデータ回復方法の一例を説明する。図 2 は、図 1 と共に述べた記憶装置システムの要部を抜き出した概略構成図である。図 2 に示す外部記憶システムは、それぞれ後述するように、ホストコンピュータ 10 と外部記憶装置とに大別され、外部記憶装置は、ディスク制御装置 200 と大容量記憶装置 400 とに大別される。ここで、図 1 と図 2 との対応関係を簡単に説明すると、図 1 中の記憶装置システム 60 が図 2 中のディスク制御装置 200

に、図 1 中のチャネル制御部 11 が図 2 中のチャネルポート 210 及びマイクロプロセッサ 220 に、図 1 中の共有メモリ 12 及びキャッシュメモリ 13 が図 2 中のバッファメモリ 230 に、図 1 中の接続部 15 がバスやスイッチ類等（図示せず）に、図 1 中のディスク制御部 14 が図 2 中のマイクロプロセッサ 220 に、図 1 中の記憶デバイス 30 が図 2 中の記憶装置 400 に、図 1 中の情報処理装置 20 が図 2 中のホストコンピュータ 100 に、それぞれ対応する。マイクロプロセッサ 220 は、チャネル制御部 11 またはディスク制御部 14 のいずれの側に存在してもよい。

【0054】

ホストコンピュータ 100 は、例えば、パーソナルコンピュータやワークステーション等から構成されるもので、データベースを扱うアプリケーションプログラム 110（以下、アプリケーションと略記）を有する。また、図示を省略しているが、ホストコンピュータ 100 は、例えば、ポインティングデバイス、キーボードスイッチ、モニタディスプレイ等を通じてオペレータと情報を交換するためのユーザインターフェースを備えている。アプリケーション 110 は、ディスク制御装置 200 を介して記憶装置 400 内のデータにアクセスすることにより、所定の業務を処理する。

【0055】

ディスク制御装置 200 は、記憶装置 400 を制御するもので、チャネルポート 210、マイクロプロセッサ 220 及びバッファメモリ 230 を備えている。

【0056】

マイクロプロセッサ 220 は、チャネルポート 210 を介して、ホストコンピュータ 100 と双方向のデータ通信を行う。マイクロプロセッサ 220 は、ディスク制御プログラム 300 を実行する。ディスク制御プログラム 300 には、書き込み制御処理 310、書き込みデータ処理 320、ディスク管理処理 330、データ回復制御処理 340、データ回復処理 350、データ同期処理 360 が含まれている。

【0057】

主要な処理については、詳細をさらに後述するが、書き込み制御処理 310 は、

主としてデータ書込み時の書込み制御情報（ジャーナル制御情報）を管理するものである。書込みデータ処理 3 2 0 は、所定のディスク装置へのデータ書込みを行うものである。ディスク管理処理 3 3 0 は、主としてジャーナルデータ格納ディスク 4 3 0 の管理を行うものである。データ回復制御処理 3 4 0 は、ホストコンピュータ 1 0 0 から設定される回復契機の登録と、登録された回復契機のリストデータをホストコンピュータ 1 0 0 に送信するものである。データ回復処理 3 5 0 は、指定されたディスク装置のデータを指定された時点まで回復させるものである。データ同期処理 3 6 0 は、ホストコンピュータ 1 0 0 からの指示に応じて、データのバックアップ処理を行うものである。

【 0 0 5 8 】

バッファメモリ 2 3 0 には、例えば、回復データ情報 D 1 0、ジャーナルデータ D 2 0、書込み制御情報 D 3 0、更新データ D 4 0 が記憶されている。回復データ情報 D 1 0 は、データの回復処理（復旧処理）の履歴情報であり、例えば、データ回復先や回復時点等を記録している。ジャーナルデータ D 2 0 は、データ操作の更新履歴であり、バッファメモリ 2 3 0 から順次ジャーナル格納ディスク 4 3 0 に移される。書込み制御情報 D 3 0 は、任意の時点にデータを回復させるために必要な情報を含んでいる。更新データ D 4 0 は、アプリケーション 1 1 0 により更新が指示されたデータであり、バッファメモリ 2 3 0 からデータ格納ディスク 4 1 0 に移される。なお、以上のデータは、バッファメモリ 2 3 0 上に同時に存在する必要はない。また、説明の便宜上、バッファメモリ 2 3 0 を単一のメモリのように示したが、例えば、複数種類のメモリ装置の集合体として構成してもよい。

【 0 0 5 9 】

大容量記憶装置 4 0 0 は、データ格納ディスク 4 1 0、バックアップデータ格納ディスク 4 2 0 及びジャーナルデータ格納ディスク 4 3 0 を備えている。データ格納ディスク 4 1 0 には、現在使用中の最新データ（実データ）が格納されている。バックアップデータ格納ディスク 4 2 0 には、ある時点のバックアップデータが格納されている。ジャーナルデータ格納ディスク 4 3 0 には、ジャーナルデータが格納されている。なお、各ディスク 4 1 0 ～ 4 3 0 は、正確にはディス

ク装置であり、それぞれ複数のディスクを備えている。以下、データ格納ディスクをデータディスク、バックアップデータ格納ディスクをバックアップディスク、ジャーナルデータ格納ディスクをジャーナルディスクと呼ぶ。

【0060】

図3は、ジャーナルデータD20及び書込み制御情報D30の概略構造を示すデータ構造図である。

【0061】

本実施の形態によるジャーナルデータD20は、書込み制御情報D30と更新データ（書込みデータ）D40とが含まれている。書込み制御情報D30は、ジャーナル制御情報としての機能を果たすもので、例えば、データ書込み位置D31、データサイズD32、タイムスタンプD33、回復フラグD34、その他制御情報D35等の情報を含んでいる。データ書込み位置D31は、どのディスク装置のどこにデータが書き込まれたかを示す位置情報である。データサイズD32は、書き込まれたデータのサイズを示す情報である。タイムスタンプD33は、データ書込み時刻を示す情報である。回復フラグD34は、回復可能な時点（復元ポイント）であることを示す標識情報であり、回復フラグD34をセットすると、回復可能なデータとして設定され、回復フラグD34をリセットすると、復元ポイントの設定が解除される。その他の制御情報D35には、例えば、書込み制御情報D30を一意に特定するための制御番号やデータ種別等のその他必要な情報が含まれる。

【0062】

本実施形態では、図3に示すように、ジャーナルデータD20の構造を独自に拡張し、ジャーナルデータD20内に回復フラグD34を設けている。これにより、少量のデータを追加するだけで任意の時点を回復可能時点として自由に設定することができ、任意の時点にデータを回復させることができる。但し、これに限らず、ジャーナルデータD20と回復フラグD34とを分離し、ユニークなID（識別コード）等で両者を対応付ける構成でもよい。

【0063】

次に、図4は、ホストコンピュータ100及びディスク制御装置200のプロ

グラム構造の概略を示すブロック図である。

【0064】

アプリケーション110は、ホストコンピュータ100のOS120を介してディスク制御プログラム300と双方向のデータ通信を行う。OS120は、API (Application Program Interface) 群130を有する。API群130には、データ書込み用API131、回復契機通知用API132、回復契機リスト取得要求用API133、回復指示用API134が含まれている。アプリケーション110は、これらAPI131～134を適宜呼び出して利用することにより、所望の時点回復契機として設定し、設定済の回復契機リストを読み出し、所望の時点を選択してデータの回復を指示することができる。

【0065】

図4を参照しつつ全体の動作を簡単に説明する。アプリケーション110が、データ書込み用API131を介して、ディスク制御装置200にデータ更新要求(コミット要求)を指示すると(S1)、ディスク制御プログラム300の書込み制御処理310は、書込みデータ処理320を介して所定のディスクにデータを書き込ませ、更新要求を処理した旨をアプリケーション110に通知する(S2)。

【0066】

アプリケーション110は、業務処理中に、例えば、定期的に又は不定期に所望の時点、回復可能な時点としての回復契機(復元ポイント)として設定することができる。アプリケーション110は、回復契機通知用API132を呼び出すことにより、回復契機を設定するデータをディスク制御装置200に対し指示する(S3)。回復契機が通知されると、ディスク制御プログラム300のデータ回復制御処理340は、指定されたデータの回復フラグをセットし、回復契機が設定された旨をアプリケーション110に通知する(S4)。

【0067】

障害発生等の要因によりデータを回復させる場合は、アプリケーション110は、回復契機リスト取得要求用API133を呼び出し、回復可能な時点のリスト情報をディスク制御装置200に要求する(S5)。リストを要求されると、

データ回復制御処理 340 は、ジャーナルディスク 430 を検査して回復フラグがセットされたデータの情報を取得し、回復契機リストを作成する。データ回復制御処理 340 は、回復契機リストをアプリケーション 110 に返信する (S6)。

【0068】

アプリケーション 110 は、メモリ 140 に格納された回復契機リストを参照し、回復を希望する時点を少なくとも 1 つ選択する。アプリケーション 110 は、回復指示用 API 134 を呼び出すことにより、希望する時点まで所定ディスクのデータを回復させるように、ディスク制御装置 200 に指示する (S8)。データ回復処理 350 は、アプリケーション 110 から回復指示を受けると、バックアップディスク 420 及びジャーナルディスク 430 を用いて、指定されたデータを指定された時点まで回復させる。回復処理 350 は、回復処理が完了した旨をアプリケーション 110 に通知する (S9)。

【0069】

次に、図 5～図 9 を参照して各部の詳細な制御を説明する。まず、図 5 は、書込み制御処理を示すフローチャートである。なお、以下の説明でも同様であるが、図示するフローチャートは発明の理解のために動作の要部を示すものであり、実際のプログラムとは相違する可能性がある。図中、「ステップ」を「S」と略記する。

【0070】

アプリケーション 110 が書込み要求を出すと、バッファメモリ 230 上のデータ D40 が更新されると共に (S21)、バッファメモリ 230 上の書込み制御情報 D30 が更新される (S22)。次に、ジャーナルディスク 430 に十分な空き容量があるか否かを判定する (S23)。例えば、ジャーナルディスク 430 の現在の空き容量が、これから書き込もうとするデータのデータサイズを上回っているか否かで判定することができる。ジャーナルディスク 430 の空き容量が不足している場合は (S23:N0)、図 6 と共に後述するジャーナルディスク管理処理を実行して空き容量を確保し (S24)、必要な場合は、バッファメモリ 230 上の書込み制御情報を更新する (S25)。必要な場合とは、例えば、後

述のジャーナル自動拡張により、ジャーナルデータの書き込み位置が変動等した場合である。

【0071】

ジャーナルディスク430に十分な空き容量が存在した場合(S23:YES)及びジャーナルディスク430に十分な空き容量が確保された場合は、書き込みデータD40及び書き込み制御情報D30(即ち、ジャーナルデータD20)をジャーナルディスク430に追加して書き込む(S26)。また、バッファメモリ230上の書き込みデータD40をデータディスク410の所定位置に書き込み(S27)、データ書き込みが完了した旨をホストコンピュータ100(正確にはホストコンピュータ100上のアプリケーション110である。以下同様)に通知する(S28)。

【0072】

なお、S26及びS27は、本書き込み制御処理とは別契機(非同期)に行っても良い。その場合、例えば、バッファメモリ上の該データにディスクへ反映したか否かのフラグを設けることにより管理することができる。

【0073】

そして、バックアップ更新フラグがオンになっているか否かを判定する(S29)。バックアップ更新フラグとは、ジャーナルディスク430の空き容量を確保するために、最古のジャーナルデータをバックアップディスク420に移し替えたことを示す標識情報である。ジャーナルデータの移し替えにより、バックアップデータから回復可能な最古の時点が変更されるため、バックアップ更新フラグがオン状態にセットされている場合は(S29:YES)、バックアップデータが更新された旨をホストコンピュータ100に通知する(S30)。バックアップ更新をホストコンピュータ100に通知した後、バックアップ更新フラグをオフ状態にリセットさせる(S31)。

【0074】

次に、図6は、図5中のジャーナルディスク管理処理S24の詳細を示すフローチャートである。まず、ジャーナルディスク430の自動拡張モードが設定されているか否かを判定する(S41)。自動拡張モードとは、未使用のディスク

、未使用の記憶領域を探索してジャーナルディスク 430 の論理的サイズを自動的に拡大させるモードである。

【0075】

自動拡張モードが設定されていない場合は (S41:NO) 、ジャーナルディスク 430 に記憶されているジャーナルデータのうち最も古いデータを選択して、バックアップディスク 420 に反映させる (S42) 。バックアップディスク 420 に移し替えられた最古のジャーナルデータは、ジャーナルディスク 430 から消去される (S43) 。これによりジャーナルディスク 430 の空き容量が増加する。ジャーナルディスク 430 の空き容量が所定値に達するまで、最古のジャーナルデータから順番にバックアップディスク 420 に移し替える (S44) 。ジャーナルディスク 430 の空き容量が所定値に達した場合は (S44:YES) 、バックアップ更新フラグをオン状態にセットする (S45) 。これにより、図 5 中の S30 に示したように、バックアップデータが更新され、バックアップデータから回復可能な最古の時点が変更された旨がホストコンピュータ 100 に通知される。なお、S44 中の所定値は、予め設定された固定値であっても良いし、例えば、バックアップディスクの空き容量やデータディスク 410 に書き込まれるデータサイズ等に応じて動的に変更される値であってもよい。

【0076】

一方、ジャーナルディスク 430 の自動拡張モードが設定されている場合は (S41:YES) 、接続されているディスク装置の中から未使用の記憶エリア (未使用エリアと呼ぶ) を検索し、ジャーナルデータを保存可能な未使用エリアが存在するか否かを判断する (S46, S47) 。未使用エリアが発見されなかった場合は (S47:NO) 、S42 に移り、上述のように最古のジャーナルデータをバックアップディスク 420 に移し替えることにより、ジャーナルディスク 430 に空き容量を確保する。未使用エリアが発見された場合は (S47:YES) 、発見された未使用エリアをジャーナルディスクとして利用すべく、ジャーナルディスク 430 の論理サイズを拡張すると共に、ディスク管理マップを更新する (S48) 。そして、ジャーナルディスク 430 の論理サイズ拡張により生じた空き容量が所定値に達したか否かを判定し (S49) 、ジャーナルディスク 430 の空き容量が

所定値に達するまで、S46～S49の処理を繰り返しながら、未使用エリアをジャーナルデータの記憶エリアとして自動的に拡張させる。

【0077】

次に、図7は、ホストコンピュータ100から指示される回復契機の登録処理を示す。上述の通り、本実施形態では、ホストコンピュータ100は、任意の時点を回復可能な契機（復元ポイント）として複数個設定することができる。

【0078】

登録すべき回復契機がホストコンピュータ100からディスク制御装置200に通知されると、データ回復制御処理340は、ジャーナルディスク430に記憶されている最新データの位置を検索し（S51）、最新の書込みデータに対応する書込み制御情報中の回復フラグをオン状態にセットして更新する（S52）。そして、ホストコンピュータ100に、回復契機の設定が完了した旨を報告すると共に、書込み制御情報を特定するための制御番号を通知する（S53）。このように、ホストコンピュータ100のアプリケーション110は、データ書込み時に、任意の時点のデータについて回復契機を設定指示できるようになっている。

【0079】

次に、図8は、ホストコンピュータ100からの要求に応じて、回復契機のリスト情報を返信する回復契機リストの送信処理を示す。まず、ジャーナルディスク430のうち、ホストコンピュータ100から回復を指定されたデータに対応するディスクを選択し、選択されたディスク中の最古のジャーナルデータにポイントを合わせる（S61）。

【0080】

そして、最古のジャーナルデータから読み込み（S62）、読み込んだジャーナルデータに関する書き込み制御情報中の回復フラグがオン状態にセットされているか否かを検査し（S63）、回復フラグがセットされている場合は、読み込んだジャーナルデータを回復契機のリスト情報に追加して記録する（S64）。S61で選択されたディスクに記憶されている最終データを読み出すまで、上記S62～S64が繰り返される（S65）。このようにして、指定されたデータ

に対応するジャーナルデータを最古のデータから最新のデータまで順番に検査し、回復フラグがセットされているジャーナルデータを抽出して回復契機リストを生成する。生成された回復契機リストは、完了報告と共にまたは非同期にホストコンピュータ100に送信される(S66)。

【0081】

次に、図9は、データ回復処理を示す。ホストコンピュータ100上のアプリケーションプログラム110は、図8に示す処理により取得した回復契機のリスト情報に基づいて、所望の時点までデータの回復を指示することができる。

【0082】

ホストコンピュータ100から回復指示が通知されると、データ回復処理350は、バックアップディスク420及びジャーナルディスク430のうち、回復が指定されたデータに対応するディスクをそれぞれ選択する(S71)。

【0083】

次に、ホストコンピュータ100からデータ回復先として指定されたディスクがバックアップディスク420であるか否かを判定する(S72)。つまり、本実施形態では、バックアップディスク420以外の他のディスク装置に、指定された時点までのデータを回復させることができるようになっている。回復先として指定されたディスク装置がバックアップディスク420以外の他のディスク装置である場合は、バックアップディスク420に記憶されているバックアップデータを、指定されたディスク装置にコピーし、データ回復の土台となるバックアップデータの準備を完了させる(S73)。

【0084】

次に、ジャーナルディスク430から最古のジャーナルデータを検索し(S74)、最古のジャーナルデータから順番にデータを読み出して、回復先に指定されたディスクの記憶内容に反映させていく(S75)。ホストコンピュータ100から指定された時点までデータが回復されるまで、ジャーナルデータを読み出して、回復先ディスクの記憶内容を更新させる(S76)。

【0085】

指定された時点までデータが回復した場合は、ホストコンピュータ100にデ

ータ回復が完了した旨を通知する（S77）。また、回復時点及び回復先等の情報を回復データ情報D10に記録する（S78）。

【0086】

本実施の形態によれば、外部記憶装置内で自動的にデータの回復を行うため、ホストコンピュータ100のコンピュータ資源をデータ回復処理のために消費することがなく、ホストコンピュータ100上の他の業務処理の効率を低下させることがない。特に、大容量の外部記憶装置を用いるアプリケーション110では、大規模なデータを取り扱うため、データ回復処理の負担が大きくなり、コンピュータ資源を多量に消費する。従って、ホストコンピュータ100上で行われる他の業務の処理速度が低下する上に、データ回復完了までの処理時間も長くなる。しかし、本実施形態では、回復契機の設定指示、回復契機リストの取得要求及び回復指示という僅かな処理だけをホストコンピュータ100で実行し、実際のデータ回復処理を外部記憶装置に委ねる構成のため、ホストコンピュータ100の負担を軽減することができる。外部記憶装置内でデータの回復を行っている間、ホストコンピュータ100は、他の業務を効率的に処理することができる。

【0087】

また、任意の複数時点を回復契機として設定可能であり、所望の時点までデータを回復可能なため、単純に直前のデータに回復させるだけの従来技術とは異なり、利便性が高い。

【0088】

さらに、本実施形態では、回復契機の設定指示や回復契機リストの取得要求等をホストコンピュータ100側から行うためのAPI131～134を用意したので、これら独自のAPIをホストコンピュータが備えるだけで、本発明に従う外部記憶装置を利用可能となる。

【0089】

また、本実施形態では、外部記憶装置内でジャーナルデータを自動的に収集すると共に、ジャーナルディスク430の空き容量管理も行うため、ジャーナルディスク430が満杯となってデータ回復が不能となるのを未然に防止できる。

【0090】

また、本実施形態では、ジャーナルデータ D20 のデータ構造を拡張し、ジャーナルデータ D20 内に（ジャーナル制御情報としての書込み制御情報 D30 内に）回復フラグを設定する構成のため、比較的簡易な構成でありながら、任意の複数時点へのデータ回復を実現することができる。

【0091】

図10は、本発明の第2の実施形態を示す。本実施形態では、複数世代のデータ管理を行っている。即ち、最新のデータを保持するデータディスク410に加えて、1世代前のデータを格納する1世代前データディスク410（1GA）、2世代前のデータを格納する2世代前データディスク410（2GA）等のように、複数世代でデータを管理できる。

【0092】

例えば、1世代前データディスク410（1GA）にバックアップディスク420の記録内容をリストアした後、ジャーナルディスク430に記憶されているデータdBのジャーナルデータを読み出して1世代前データディスク410（1GA）に反映させれば、1世代前のデータに戻すことができる。同様に、2世代前データディスク410（2GA）に、バックアップデータをコピーしてからデータdB及びデータdCのジャーナルデータを反映させることにより、2世代前のデータに戻すことができる。このように、複数世代でデータを管理する場合も、本発明に従えば、ホストコンピュータ100に処理負担をかけることなく、外部記憶装置内で複数世代のデータを構築し管理することができる。

【0093】

なお、本発明は、上述した各実施の形態に限定されない。当業者であれば、本発明の範囲内で、種々の追加や変更等を行うことができる。

【図面の簡単な説明】

【図1】

本発明の第1の実施の形態に係る外部記憶システムの概略構成図である。

【図2】

図1に示す記憶装置システムの概略を示すブロック図である。

【図3】

ジャーナルデータ及び書込み制御情報の構造を示すデータ構造図である。

【図4】

ホストコンピュータ及びディスク制御装置のプログラム構造を示すブロック図である。

【図5】

書込み制御処理を示すフローチャートである。

【図6】

ジャーナルディスク管理処理を示すフローチャートである。

【図7】

ホストコンピュータから回復契機を通知された場合のデータ回復制御処理を示すフローチャートである。

【図8】

ホストコンピュータから回復契機リストの送信を要求された場合のデータ回復制御処理を示すフローチャートである。

【図9】

ホストコンピュータから回復を指示された場合のデータ回復処理を示すフローチャートである。

【図10】

複数世代でデータ管理を行う場合の模式図である。

【符号の説明】

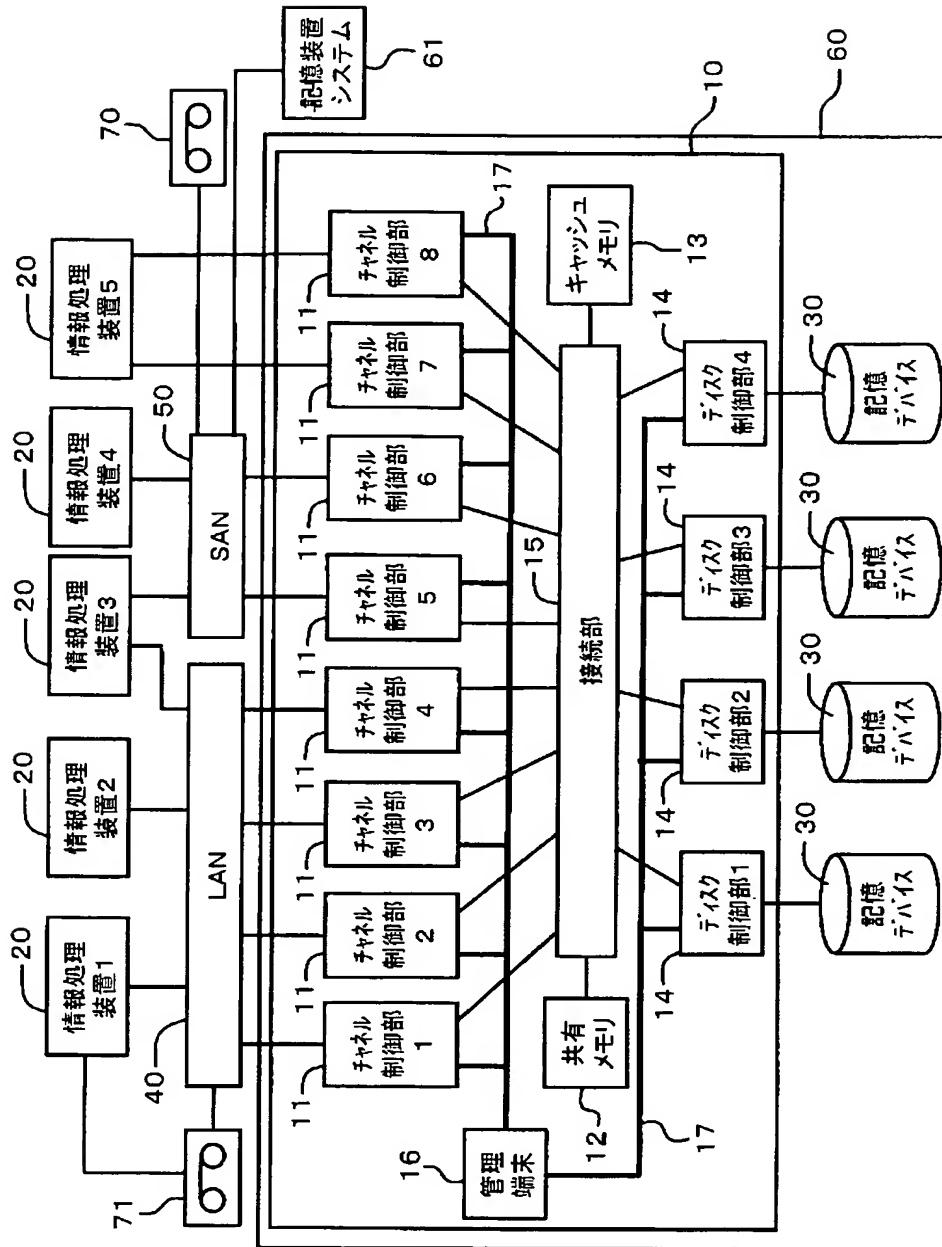
- 10 記憶デバイス制御装置
- 11 チャンネル制御部
- 12 共有メモリ
- 13 キャッシュメモリ
- 14 ディスク制御部
- 15 接続部
- 16 管理端末
- 17 内部LAN
- 20 情報処理装置

- 30 記憶デバイス
- 40 LAN
- 50 SAN
- 60, 61 記憶装置システム
- 70, 71 バックアップデバイス
- 100 ホストコンピュータ
- 110 アプリケーションプログラム
- 120 OS
- 130 API
- 131 データ書込み用API
- 132 回復契機通知用API
- 133 回復契機取得要求用API
- 134 回復指示用API
- 140 メモリ
- 200 ディスク制御装置
- 210 チャネルポート
- 220 マイクロプロセッサ
- 230 バッファメモリ
- 300 ディスク制御プログラム
- 310 書込み制御処理
- 320 書込みデータ処理
- 330 ディスク管理処理
- 340 データ回復制御処理
- 350 データ回復処理
- 360 データ同期処理
- 400 大容量記憶装置
- 410 データ格納ディスク装置
- 420 バックアップデータ格納ディスク装置
- 430 ジャーナルデータ格納ディスク装置

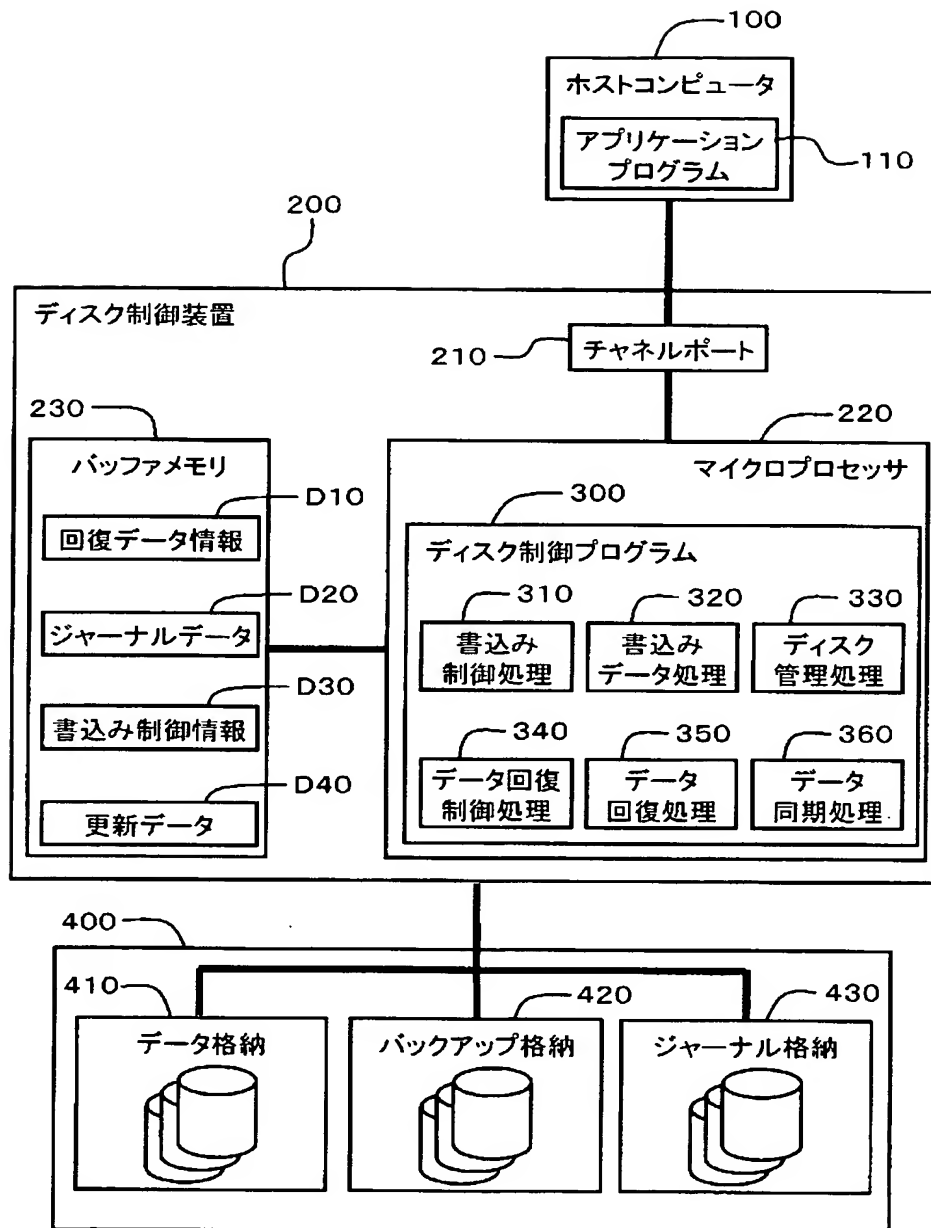
【書類名】

図面

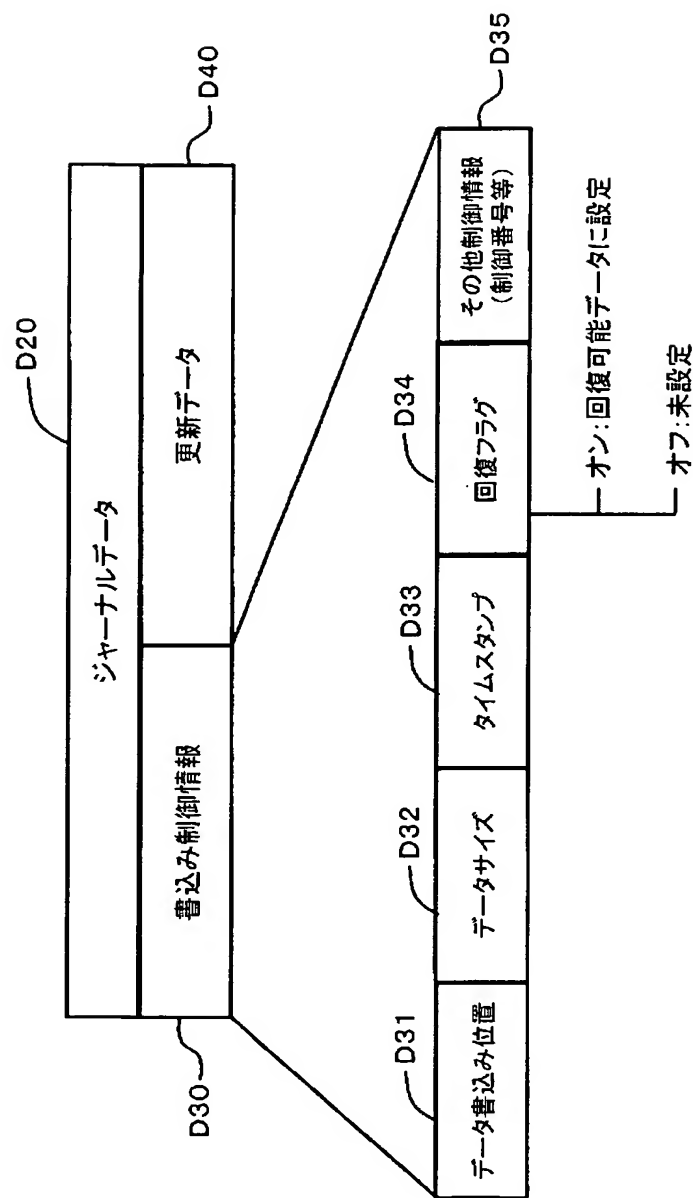
【図 1】



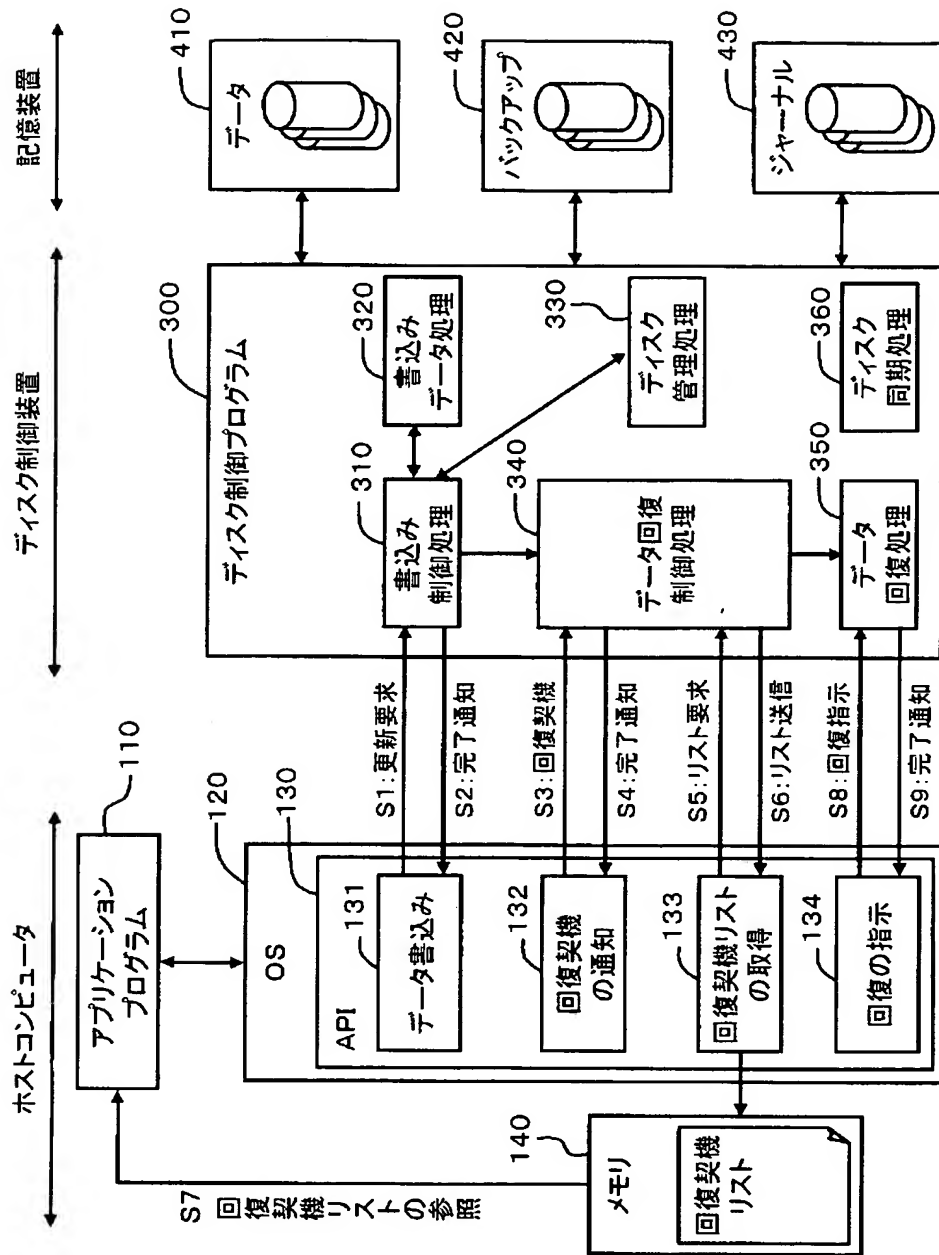
【図 2】



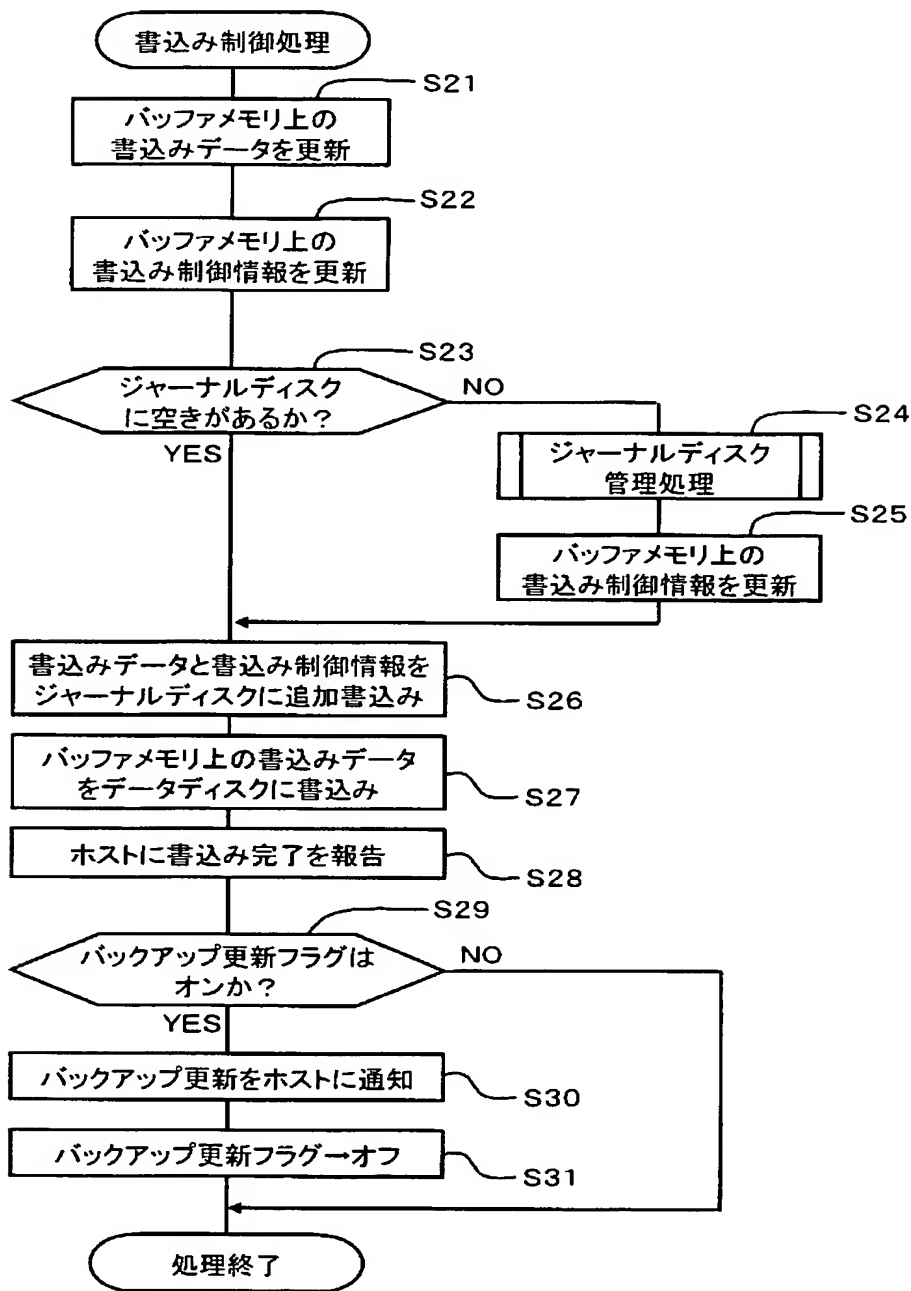
【図 3】



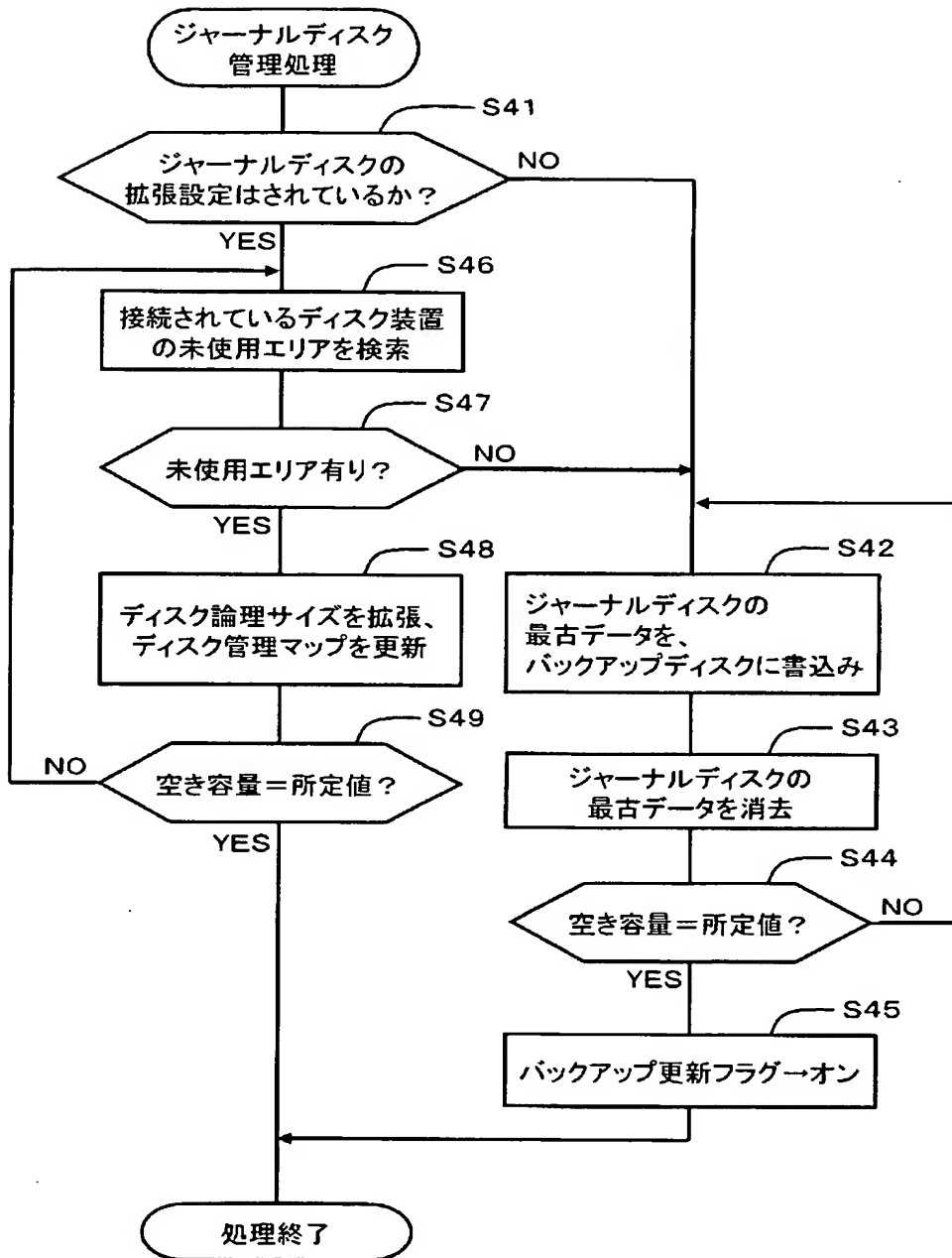
【図 4】



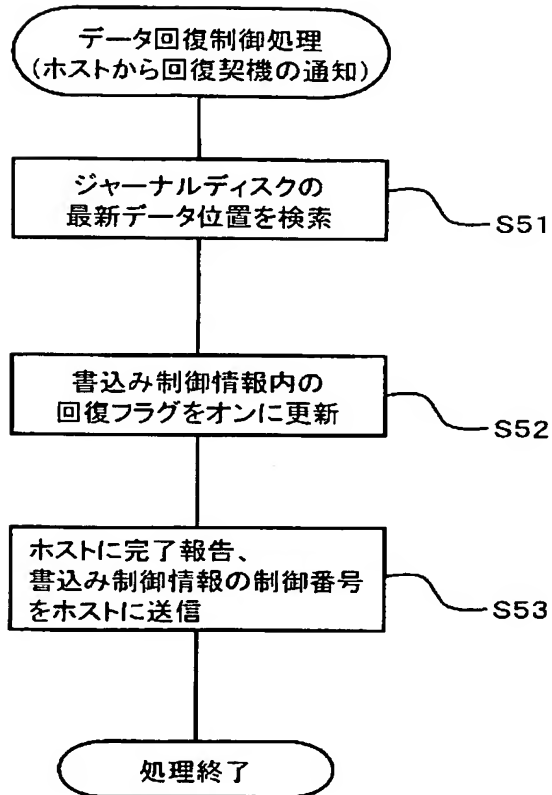
【図 5】



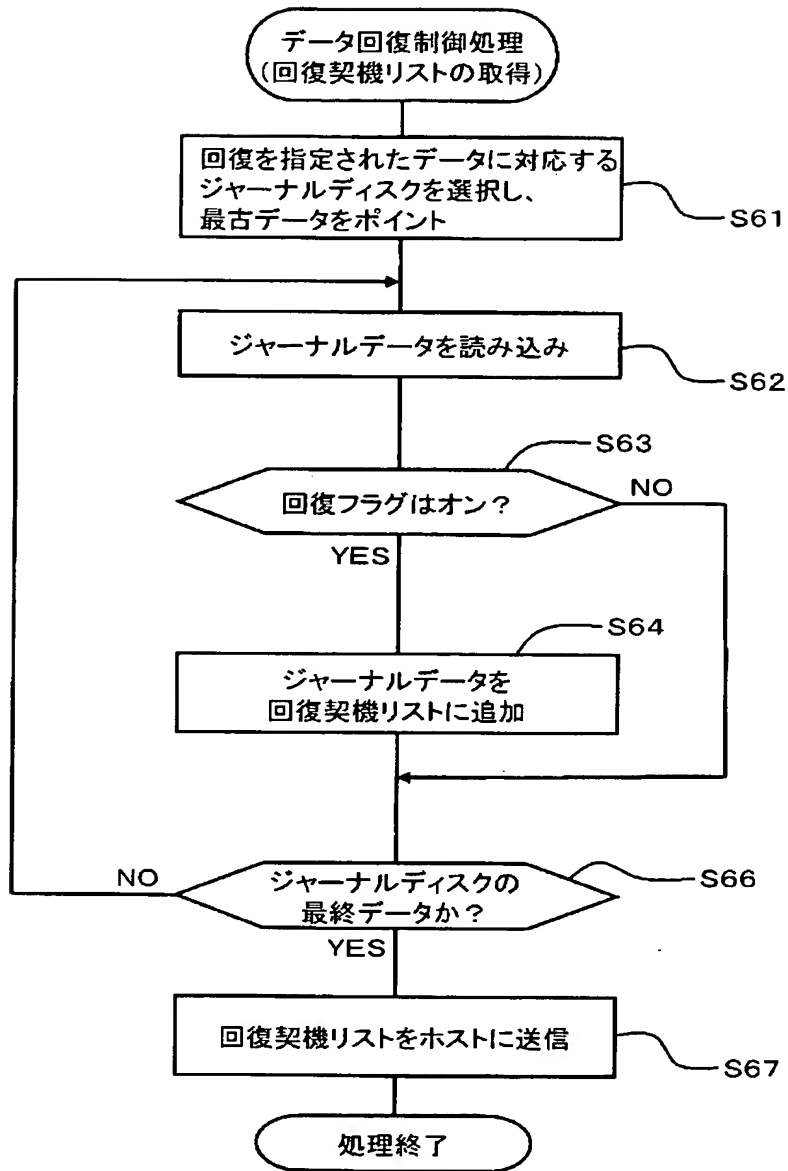
【図 6】



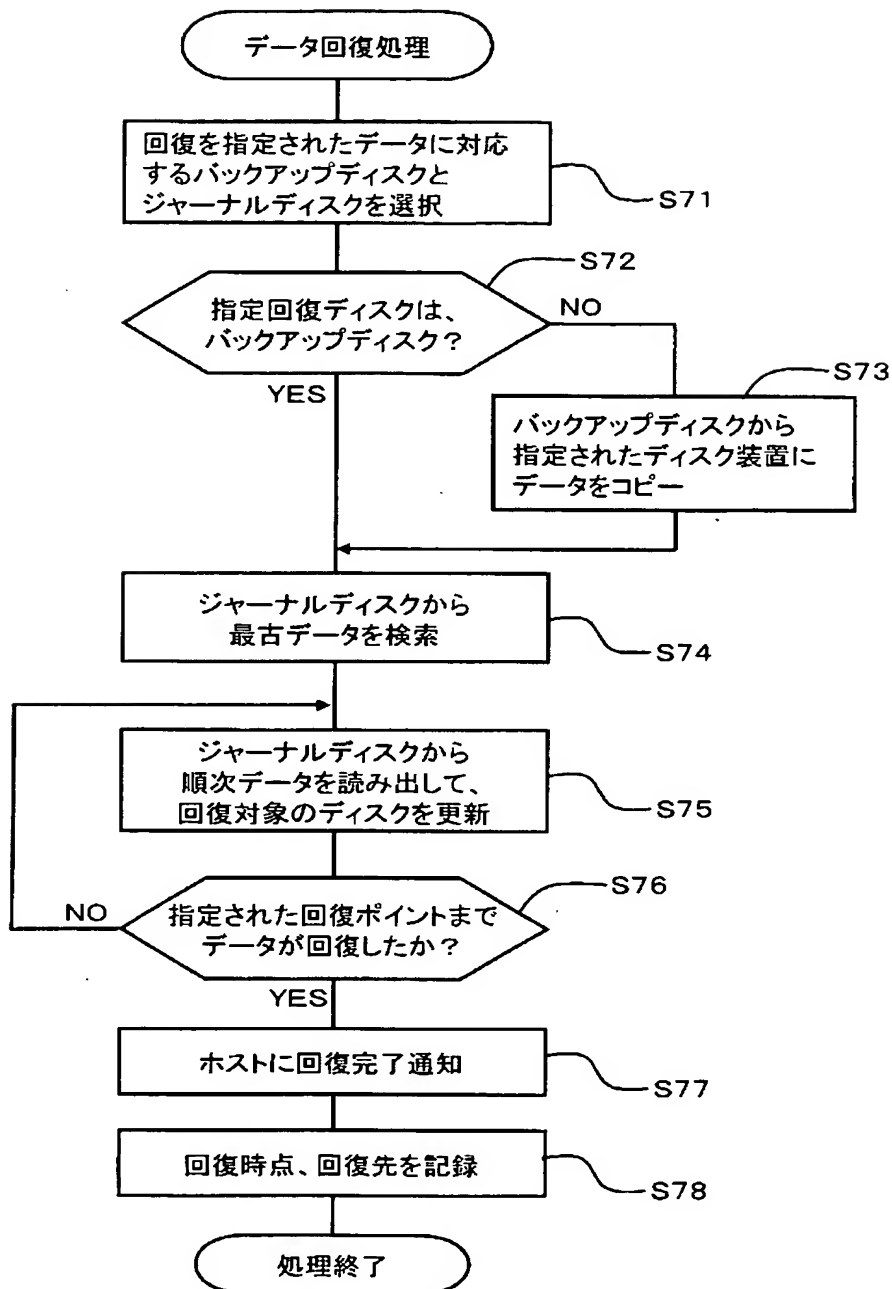
【図 7】



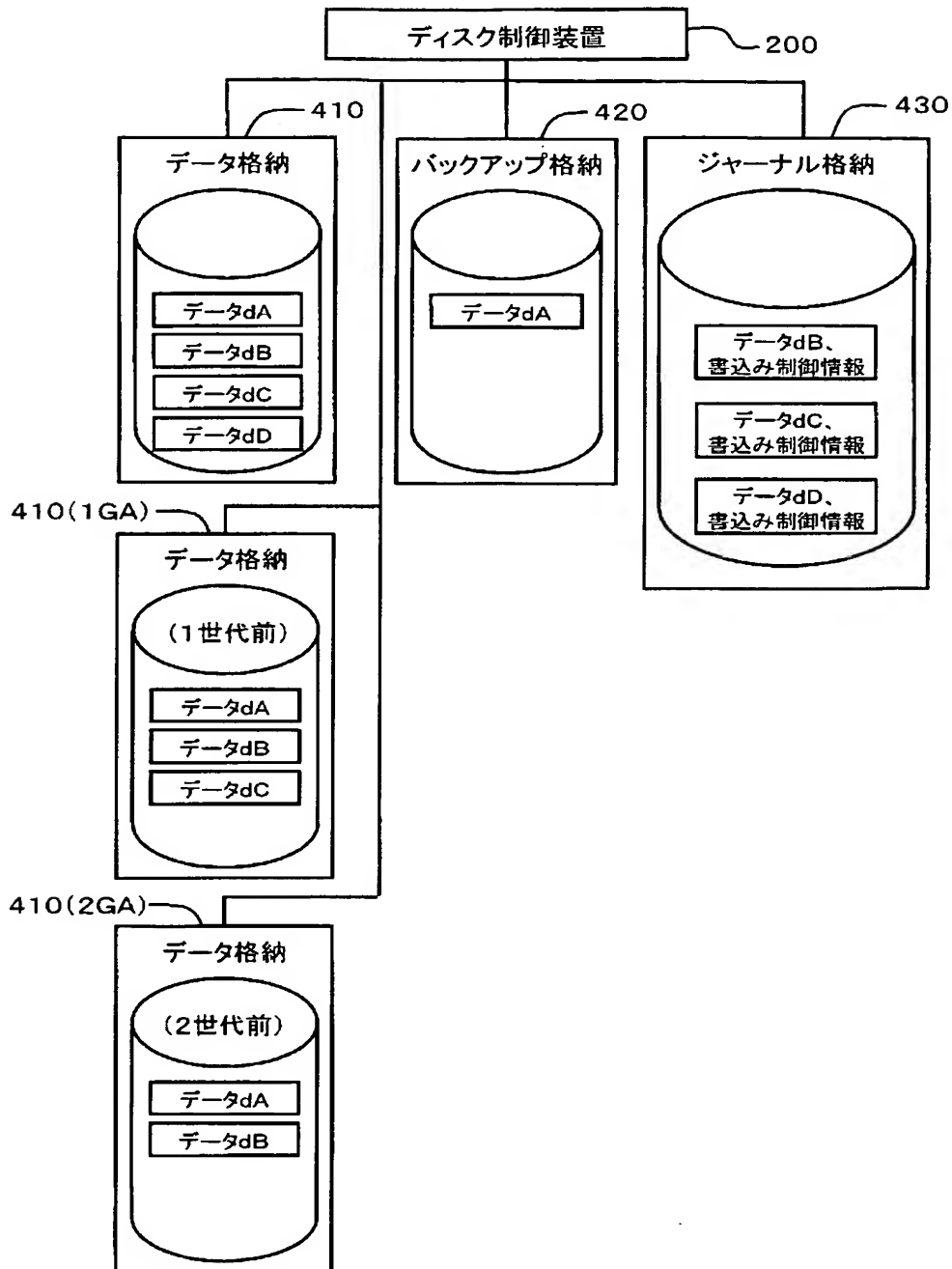
【図 8】



【図 9】



【図 10】



【書類名】 要約書

【要約】

【課題】 ホストコンピュータに負担をかけずに、外部記憶装置内で、所望の任意の時点までデータを自動的に回復させる。

【解決手段】 ホストコンピュータ上のアプリケーション 110 は、ディスク制御装置のデータ回復制御処理 340 に対し、回復契機の設定を指示する（S3）。ジャーナルデータ中に含ませた回復フラグをセットすることにより、任意の複数時点を回復可能な時点として登録させることが可能となっている。障害等が発生してデータを回復させる場合、アプリケーション 110 は、設定済の回復契機の一覧を示すリストを要求する（S5）。アプリケーション 110 は、回復契機リストに基づいて、データを回復させる時点を指定する（S8）。ディスク制御装置は、バックアップディスク 420 及びジャーナルディスク 430 に基づいて、指示された時点までデータを回復させる。

【選択図】 図 4

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 0 7 6 8 6 5
受付番号	5 0 3 0 0 4 5 5 3 1 7
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 3 月 2 4 日

< 認定情報・付加情報 >

【提出日】 平成15年 3月20日

次頁無

特願 2 0 0 3 - 0 7 6 8 6 5

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所